12

EDGEWOOD RESEARCH, DEVELOPMENT & ENGINEERING CENTER

ERDEC-TR-091

SYMMETRIC ORIENTATIONS FOR SIMPLEX DESIGNS

Ronald B. Crosier

RESEARCH AND TECHNOLOGY DIRECTORATE

July 1993

U.S. ARMY CHEMICAL AND BIOLOGICAL DEFENSE AGENCY

Aberdeen Proving Ground, Maryland 21010-5423

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
1993 July

3. REPORT TYPE AND DATES COVERED
Final, 92 Aug - 93 Mar

4. TITLE AND SUBTITLE
Symmetric Orientations for Simplex Designs

5. FUNDING NUMBERS
PR-10161102A71A

6. AUTHOR(S)
Crosier, Ronald B.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DIR, ERDEC,* ATTN: SCBRD-RTM, APG, MD 21010-5423

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDEC-TR-091

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
*When this work was performed, ERDEC was known as the U.S. Army Chemical Research, Development and Engineering Center, and the author was assigned to the Research Directorate.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Designs for fitting first-order response surfaces over spherical regions are often constructed from the vertices of regular geometric figures. Designs constructed from cubes (the two-level factorial designs) are naturally oriented so that the designs treat the factors equally, but the simplex designs are usually presented in an orientation that gives the factors unequal, asymmetric ranges. The unequal, asymmetric ranges of the coded design factors create a problem in applying the designs. The geometric and statistical properties of the design must be distorted to give the experimental factors the desired ranges. Symmetric orientations are given for the simplex designs for 4, 5, 6, 8, 9, 10, 12, and 13 factors. The simplex designs for 5 or more factors have 6 levels of the factors in the new orientations.

14. SUBJECT TERMS
Design rotation   Response surface design

15. NUMBER OF PAGES
13

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Blank

PREFACE

The work described in this report was authorized under Project No. 10161102A71A, Research in CW/CB Defense. This work was started in August 1992 and completed in March 1993.

The use of trade names or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for release to the public. Registered users should request additional copies from the Defense Technical Information Center; unregistered users should direct such requests to the National Technical Information Service.

Acknowledgments

The author wishes to acknowledge James O. Jensen, U.S. Army Edgewood Research, Development and Engineering Center (ERDEC),* for technical assistance on this project.

* When this work was performed, ERDEC was known as the U.S. Army Chemical Research, Development and Engineering Center, and the author was assigned to the Research Directorate.

Blank

CONTENTS

Blank

SYMMETRIC ORIENTATIONS FOR SIMPLEX DESIGNS

1. INTRODUCTION

In many industrial experiments the goal is to determine which process variables or factors $x_1, x_2, \ldots, x_k$ affect the yield y of the product, or influence some characteristic of the product. A common approach to such problems is to approximate the relationship between y and the process factors by a first-order polynomial,

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \epsilon. \tag{1}$$

The coefficients of the polynomial are estimated from data collected during N experimental runs of the process; the settings of the x's for the N experimental runs are given by a response surface design. A response surface design for k factors is written as an $N \times k$ design matrix $\mathbf{D}$. To estimate the coefficients of the polynomial, a column of 1's to represent the intercept term is appended to the design matrix and the expanded design matrix $\mathbf{X}$ is used to estimate the coefficient vector β by the least-squares formula

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{2}$$

where $\mathbf{b}$ is a vector of estimates for the coefficients and $\mathbf{y}$ is an $N \times 1$ vector of responses. Tests of statistical significance of the coefficients in (1) are used to determine which factors have an effect that clearly stands out against the random variation (background noise) in the data. Another type of analysis is to use the estimated coefficients to rank the process factors from most important to least important, without doing statistical tests.

Biles and Swain (1980) and Box and Draper (1987) describe the use of simplex designs for fitting the first-order model (1) over a spherical region. A regular simplex is the geometric figure formed by $k+1$ equally spaced points in k-dimensional space; an equilateral triangle is a regular simplex in two dimensions. In this report, simplex design will mean a design whose design points are the vertices of a regular simplex. Among the advantages of the simplex design are that it requires the minimum number of runs to estimate the parameters in the model (1) and that it estimates the parameters with maximum precision. To obtain estimates of lack of fit and pure error, n_0 center points can be added to the simplex design. The expanded design matrix $\mathbf{X}$ will then have another column to test for lack of fit. The lack-of-fit column should be made orthogonal to the other columns of $\mathbf{X}$ by making the value in the lack-of-fit column for the simplex points proportional to n_0 and the value in the lack-of-fit column for the center points proportional to $-(k+1)$.

The statistical properties of a simplex design (orthogonal, minimum variance estimates of the coefficients and the lack-of-fit test using center points) are not affected by the orientation of the simplex relative to the coordinate axes. Two orientations for simplex designs are commonly discussed in the literature. One orientation allows a simplex design to be developed for any number of factors. The columns of the design matrix are obtained by scaling a set of simple contrasts to have the same sum of squares. For $j=1, 2, 3, \ldots, k$, the jth contrast (column of $\mathbf{D}$) before scaling consists of j ones, $-j$, and zeros for the remaining values. I call this orientation of the simplex designs the Helmert orientation because the $(k+1)\times(k+1)$ expanded design matrix $\mathbf{X}$, when scaled so that $\mathbf{X}'\mathbf{X}=\mathbf{I}$, is the orthogonal matrix that represents Helmert's transformation [see, for example, Kendall (1961, page 12)].

The other widely discussed orientation of simplex designs yields two-level designs for $k=3, 7, 11, \ldots$. These designs are best known to statisticians as the Plackett and Burman (1946) designs. When scaled so that $\mathbf{X}'\mathbf{X}=(k+1)\mathbf{I}$, the expanded design matrix $\mathbf{X}$ is an Hadamard matrix of order $k+1$, so I refer to this orientation of the simplex designs as the Hadamard orientation. These two-level designs are popular, but they are only available for $k=3, 7, 11, \ldots$.

2. PROBLEMS WITH THE HELMERT ORIENTATION

Table 1 gives the four-factor simplex design, scaled so that $\mathbf{X}'\mathbf{X}=(k+1)\mathbf{I}$, in the Helmert orientation.

Table 1. Four-Factor Simplex Design In Helmert Orientation

Point	x_1	x_2	x_3	x_4
1	1.581	.913	.645	.5
2	−1.581	.913	.645	.5
3	0.	−1.826	.645	.5
4	0.	0.	−1.936	.5
5	0.	0.	0.	−2.

Consider the application of the design in Table 1 to the experimental factors blender speed (100-300 revolutions per minute, or rpm), mixing time (10-20 minutes), cooking temperature (180-200 degrees), and cooking time (40-60 minutes). Because the design points in Table 1 lie on a sphere of radius 2 in coded units, a method of applying the design in Table 1 is to associate the design diameter of 4 coded units with the range of the experimental factors. Thus the coded range (−2,2) corresponds to the range of the experimental factors. Because the first factor in Table 1 has a coded range of (−1.581,1.581), blender speed would vary over the interval 121-279 rpm rather than the desired 100-300 rpm. By this method of applying the design, mixing time varies from 10.4 minutes to 17.3 minutes, cooking temperature from 180.3° to 193.2°, and cooking time from 40 minutes to 52.5 minutes. Although this method of applying the design is consistent with the theory of optimal designs, the reduced and asymmetric ranges of the experimental factors are generally unacceptable to experimenters.

An alternative is to apply the design by scaling the coded factor ranges to the ranges of the experimental factors, but then the design is no longer a regular simplex. Box and Draper (1987) mention this point, but do not elaborate. To see the disadvantages of this method of applying the design, examine the levels of the experimental factors. For blender speed, the coded levels −1.581, 0, and 1.581 become 100, 200, and 300 rpm; the coded levels −1.826, 0, and .913 of the second factor become 10, 16.7, and 20 minutes for mixing time. Notice that the asymmetric range of the second design factor has produced a central level (16.7 minutes) that is not the midpoint of the range (10-20 minutes). For cooking time, the coded levels −2 and .5 become 40 and 60 minutes, respectively. Now consider a modified design obtained by replacing the coded levels −2 and .5 of the last factor by the coded levels −1 and 1; the last column in Table 1 would then have a 1 in rows 1-4 and a −1 in row 5. Applying the modified design by scaling the coded factor ranges to the ranges of the experimental factors yields the same design for the experimental factors as the unmodified design. One possible conclusion is that this method of applying a response surface design is simply wrong. Another conclusion is that each column of the coded design matrix can be rescaled without affecting the levels of the experimental factors. It can be argued that this method of applying a response surface design is equivalent to rescaling the columns of the coded design matrix to cover the same interval. But such a rescaling would change the geometric and statistical properties of the design. The Helmert orientation of the simplex designs therefore gives the practitioner the choice of a design without the desired ranges for the experimental factors or a design whose statistical properties have been distorted by its method of application.

An alternative to the Helmert orientation of the simplex designs, discussed by Box and Draper (1987), is the use of the next larger Hadamard orientation simplex design. Thus one would use the seven-factor design for four factors by ignoring three of the design columns. The disadvantage of using the seven-factor design for four factors is that the seven-factor design requires three more experimental runs than the four-factor design.

3. A SYMMETRIC ORIENTATION FOR THE FOUR-FACTOR DESIGN

Table 2 gives the four-factor simplex design in a symmetric orientation—meaning that the coded design factors have the same range and the common range is symmetric about zero.

Table 2. Four-Factor Simplex Design In New Orientation

Point	x_1	x_2	x_3	x_4
1	.309	.691	1.309	−1.309
2	.691	1.309	−1.309	.309
3	1.309	−1.309	.309	.691
4	−1.309	.309	.691	1.309
5	−1.	−1.	−1.	−1.

All the coded factors cover the range −1.309 to 1.309, and the design can be applied by scaling the coded factor range to the ranges of the experimental factors without the difficulties created by the Helmert orientation. Note that the design consist of the cyclic permutations of .309, .691, 1.309, and −1.309 appended by a row of −1's and is scaled so that $\mathbf{X}'\mathbf{X} = (k+1)\mathbf{I}$. This construction and scaling was used by

Plackett and Burman (1946) for most of the simplex designs in the Hadamard orientation.

4. SYMMETRIC ORIENTATIONS FOR OTHER SIMPLEX DESIGNS

For two factors, there is no symmetric orientation of the simplex design: either the factors have different ranges, say, $-a$ to a and $-b$ to b, or the common range is asymmetric about zero, say, $-a$ to b. For $k=3$ to $k=13$, symmetric orientations of simplex designs can be generated from the numeric values in Table 3 by using the cyclic permutations of the values and appending a row of -1's. The designs for $k=3$, 7, and 11 are (equivalent to) the Plackett and Burman (1946) designs. Notice that the numeric values for odd k have a 1 in the middle position and the values above and below the middle position are equal in magnitude but opposite in sign. The designs for odd k therefore have the property that if a factor occurs at level c, it also occurs at level $-c$ (the 1 in the middle position is balanced by the -1 in the appended row of -1's). The only disadvantage of the new symmetric orientations is that the designs require more levels than the Hadamard orientation. Over the range of k in Table 3 the number of levels required does not exceed six (seven if center points are used). The new orientations therefore allow a choice between minimizing the number of runs or minimizing the number of levels by going to the next larger Hadamard design. In some applications, the number of levels may not be relevant, or many levels of the factors may even be desirable [see Welch, Buck, Sacks, Wynn, Mitchell, and Morris (1992)]. The correctness of the designs can be verified by showing that $\mathbf{X}'\mathbf{X} = (k+1)\mathbf{I}$.

Table 3. Numeric Values for Simplex Designs. The cyclic permutations of the k values below, appended by a row of -1's, form the k-factor regular simplex design.

	k										
	3	4	5	6	7	8	9	10	11	12	13
	1	.309	.325	−.274	1	.25	.577	−.232	1	.217	.667
	1	.691	−1.376	1.274	−1	.75	1.732	1.232	1	.783	.667
	−1	1.309	1.000	−1.274	−1	1.25	−.577	−1.232	1	−.783	−1.247
		−1.309	1.376	−.726	1	.75	.577	−.768	−1	−1.217	1.247
			−.325	.726	1	−.75	1.000	−1.232	1	1.217	.667
				1.274	1	−1.25	−.577	1.232	1	−1.217	1.247
					−1	1.25	.577	.768	−1	1.217	1.000
						−1.25	−1.732	−.768	1	.783	−1.247
							−.577	.768	−1	1.217	−.667
								1.232	−1	.783	−1.247
									−1	−.783	1.247
										−1.217	−.667
											−.667
Levels:	2	5	6	6	2	6	6	6	2	6	6

NOTE: For $k=3$, 7, 8, and 11, the values are exact. For even k, the numeric values are $u=((k+1)^{1/2}-1)/k$, or $-u$, and $\pm 1 \pm u$. For $k=5$ and 13, the values are $\pm(1 \pm 2/k^{1/2})^{1/2}$, and 1. For $k=9$, the values are $\pm 1/3^{1/2}$, 1, and $\pm 3^{1/2}$.

LITERATURE CITED

Biles, W. E. and Swain, J. J. (1980), *Optimization And Industrial Experimentation*, John Wiley, New York.

Box, G. E. P. and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, John Wiley, New York.

Kendall, M. G. (1961), *A Course in the Geometry of n Dimensions*, Hafner Publishing Company, New York.

Plackett, R. L. and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika*, 33, 305-325.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992), "Screening, Predicting, and Computer Experiments," *Technometrics*, 34, 15-25.